

AD-A044 909

STANFORD UNIV CALIF SYSTEMS OPTIMIZATION LAB
FUNDAMENTALS OF QUADRATIC PROGRAMMING AND LINEAR COMPLEMENTARITY--ETC(U)

UNCLASSIFIED

AUG 77 R W COTTLE

SOL-77-21

F/G 12/1

F44620-74-C-0079

NL

| OF |
AD:
A044 909



END
DATE
FILMED
10-77
DDC

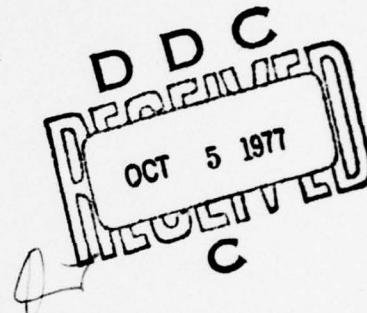
AD A 044 909

FUNDAMENTALS OF QUADRATIC PROGRAMMING AND LINEAR COMPLEMENTARITY

BY

RICHARD W. COTTLE

TECHNICAL REPORT SOL 77-21
AUGUST 1977



Systems Optimization Laboratory

Department of
Operations
Research

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

Stanford
University

Stanford
California
94305

AD No.
DDC FILE COPY

FUNDAMENTALS OF QUADRATIC PROGRAMMING AND LINEAR COMPLEMENTARITY*

by

Richard W. Cottle

TECHNICAL REPORT SOL 77-21

August 1977



SYSTEMS OPTIMIZATION LABORATORY
DEPARTMENT OF OPERATIONS RESEARCH

Stanford University
Stanford, California

*This paper was presented at the NATO Advanced Study Institute on
Engineering Plasticity by Mathematical Programming, University
of Waterloo, Waterloo, Ontario, Canada, August 2-12, 1977.

Research and reproduction of this report were partially sponsored by the
Air Force Office of Scientific Research Contract F-44620-74-C-0079 and
the Office of Naval Research Contract N-00014-75-C-0267.

Reproduction in whole or in part is permitted for any purpose of the
United States Government. This document has been approved for public
release and sale; its distribution is unlimited.

FUNDAMENTALS OF QUADRATIC PROGRAMMING AND LINEAR COMPLEMENTARITY

Richard W. Cottle

*Department of Operations Research
Stanford University, Stanford, California*

1. INTRODUCTION

Quadratic programming and linear complementarity are two closely related mathematical programming specialties having strong relevance in engineering plasticity, as well as other seemingly distant disciplines. The aim of this paper is to provide a coherent introduction to these subjects. The presentation is deliberately focussed on fundamentals, the purpose being to communicate the basic theoretical and algorithmic concepts and to indicate where the reader may turn for more details.

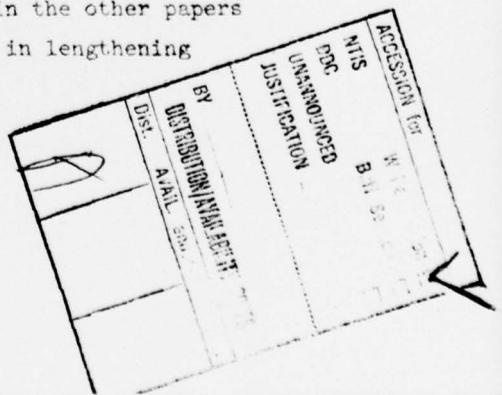
The discussion of quadratic programming given here is essentially limited to the convex case. Fortunately, this covers most of the commonly encountered applications of QP. The development includes a presentation of the optimality conditions and the construction of dual quadratic programs. A few of the many available convex quadratic programming algorithms are also described.

The consideration of optimality conditions in QP gives rise to the formulation of the linear complementarity problem. The LCP is a useful framework within which the QP and its associated algorithms can be studied. However, the LCP arises in other ways, too. Some highlights of the theory and a selection of the principal computational techniques are assembled in this paper.

In some applications, it is necessary to solve a parametric QP or LCP. Such problems are discussed here from the theoretical and computational standpoints.

Potential users of QP and LCP software will definitely be interested in what can be expected in terms of computational efficiency. The available information on this is somewhat scanty. Nevertheless, an attempt has been made to summarize what has been reported in the literature.

For reasons of space, the proofs of the theorems quoted here are omitted, but they can be found in the cited references. As for specific engineering plasticity applications, many will be found in the other papers of these Proceedings, so there would seem to be no point in lengthening this article by developing them again.



2. THE FORMULATION AND GEOMETRY OF QUADRATIC PROGRAMMING

Quadratic programming is that part of mathematical programming concerned with the minimization (or maximization) of a quadratic function of many variables subject to linear constraints. As a rule, these linear constraints include inequalities. Indeed, from the standpoint of theory, it is possible to convert any linear constraints to linear inequalities in nonnegative variables. Throughout this paper, the statement of the quadratic programming problem will be

$$\begin{aligned} & \text{minimize } \tilde{c} \underline{x} + \frac{1}{2} \tilde{x} \underline{H} \underline{x} \\ & \text{subject to } \underline{A} \underline{x} \leq \underline{b} \\ & \quad \underline{x} \geq \underline{0} \end{aligned} \tag{1}$$

In this formulation,

$$\underline{A} \in \mathbb{R}^{m \times n}, \underline{b} \in \mathbb{R}^m, \underline{c} \in \mathbb{R}^n, \underline{H} \in \mathbb{R}^{n \times n}$$

are given; the quadratic function

$$\Phi(\underline{x}) = \tilde{c} \underline{x} + \frac{1}{2} \tilde{x} \underline{H} \underline{x}$$

is the objective function, and the set

$$\underline{C} = \{ \underline{x} \in \mathbb{R}^n \mid \underline{A} \underline{x} \leq \underline{b}, \underline{x} \geq \underline{0} \}$$

is the constraint set (or feasible set). One can assume that the matrix \underline{H} appearing in the definition of Φ is symmetric, for otherwise, it can always be replaced by its symmetric part, $\frac{1}{2} (\underline{H} + \tilde{\underline{H}})$.

The constraint set \underline{C} is always convex and polyhedral, whereas the objective function Φ is convex on \mathbb{R}^n if and only if \underline{H} is positive semi-definite, meaning

$$\tilde{x} \underline{H} \underline{x} \geq 0 \quad \text{for all } \underline{x} \in \mathbb{R}^n. \tag{2}$$

When this condition obtains, the QP (1) is called a convex quadratic program. If \underline{H} is positive definite, so that (2) holds as a strict inequality for all $\underline{x} \neq \underline{0}$, the function Φ is strictly convex, and (1) is then called a strictly convex quadratic program. When Φ is not convex on \mathbb{R}^n , its restriction to \underline{C} can still be convex if that set is of lower dimension than n . At any rate, when (2) is not valid, (1) is commonly called a nonconvex QP.

Among QP problems, the convex case is much easier to deal with both theoretically and methodologically. Fortunately, the applications of convex QP are sufficiently widespread to make it worthy of independent development. In the author's opinion, it is of greater importance to be able to solve large convex quadratic programming problems than nonconvex problems of small size. Accordingly, the convexity assumption will be rather freely invoked throughout this survey.

A vector $\underline{x}^* \in \underline{C}$ such that $\varphi(\underline{x}^*) \leq \varphi(\underline{x})$ for all $\underline{x} \in \underline{C}$ is said to be optimal for (1).

The constraints of a QP are like those of an LP, and when $\underline{H} = \underline{0}$, (1) really is an LP. It is a well known fact that if an LP has an optimal solution, then it must have an optimal solution which is also an extreme point of \underline{C} . In this respect, a convex QP is unlike an LP. It is not necessary for its optimal solutions, if any, to occur at extreme points of \underline{C} . Simple illustrations can be given to illustrate that there exist convex QP problems with interior point solutions, boundary point solutions, extreme point solutions or no solutions. Despite this major difference between LP and QP, there are elements of great similarity between the two classes of problems. Some of these will be emphasized in the next two sections.

3. OPTIMALITY CONDITIONS AND DUALITY

3.1. Optimality Conditions. In all branches of mathematical programming, it is essential to know how to recognize an optimal solution. In QP, this is done with the Kuhn-Tucker conditions [32], [31] which were developed for and apply to nonlinear programs of much greater generality. The presentation of optimality conditions given in this section is a specification of "Kuhn-Tucker theory" to the QP case.

The following result expresses the necessary conditions of optimality for a QP of the form (1). The theorem is valid regardless of whether the objective function is convex.

Theorem 1. If \underline{x}^* is an optimal solution to the QP (1), then there exists a vector \underline{y}^* such that

$$\underline{c} + \underline{H} \underline{x}^* + \tilde{\underline{A}} \underline{y}^* \geq \underline{0} \quad (3)$$

$$\tilde{\underline{A}}^* (\underline{c} + \underline{H} \underline{x}^* + \tilde{\underline{A}} \underline{y}^*) = \underline{0} \quad (4)$$

$$\tilde{\underline{y}}^* (\underline{b} - \underline{A} \underline{x}^*) = \underline{0} \quad (5)$$

$$\underline{y}^* \geq \underline{0} \quad (6)$$

Equations (3) - (6) are called the Kuhn-Tucker conditions for (1). There are several points worth noting about them. First, conditions (4) and (5) are known as the complementary slackness conditions. They have the following significance. Since

$$\underline{y}^* = \underline{c} + \underline{H} \underline{x}^* + \tilde{\underline{A}} \underline{y}^* \geq \underline{0}$$

and $\underline{x}^* \geq \underline{0}$, the equation (4) implies that

$$x_j^* y_j^* = 0 \quad j = 1, \dots, n \quad (7)$$

More emphatically

$$\begin{aligned} x_j^* > 0 \text{ implies } u_j^* = 0 & \quad j = 1, \dots, n \\ u_j^* > 0 \text{ implies } x_j^* = 0 \end{aligned} \quad (8)$$

Since \underline{x}^* must be feasible,

$$\underline{y}^* = \underline{b} - \underline{A} \underline{x}^* \geq \underline{0}.$$

Thus (5) says

$$y_i^* v_i^* = 0 \quad i = 1, \dots, m \quad (9)$$

and

$$\begin{aligned} y_i^* > 0 \text{ implies } v_i^* = 0 & \quad i = 1, \dots, m \\ v_i^* > 0 \text{ implies } y_i^* = 0 \end{aligned} \quad (10)$$

It does not follow, however, that the implications (8) and (10) can be reversed.

When $v_i = (\underline{b} - \underline{A} \underline{x})_i = 0$, the i -th constraint is said to be binding (or active) at \underline{x} . Notice that if $\underline{x}^* \geq \underline{0}$ is optimal for (1) and no constraints are binding at \underline{x}^* , then $\underline{u}^* = \underline{0}$ and $\underline{y}^* = \underline{0}$, so that \underline{x}^* is an interior point of \underline{C} , and the optimality condition asserts the familiar equation $\nabla \varphi(\underline{x}^*) = \underline{c} + \underline{H} \underline{x}^* = \underline{0}$.

The components of \underline{y}^* are known as Lagrange multipliers for the corresponding constraints $\underline{b} - \underline{A} \underline{x} \geq \underline{0}$. The components of \underline{u}^* are Lagrange multipliers for the constraints $\underline{x} \geq \underline{0}$. The numbers y_i^* can be interpreted as measures of the rate of change of the minimum value of the objective function φ with respect to changes in the corresponding components of \underline{b} . This interpretation is worked out in [53], [14].

The Kuhn-Tucker conditions are merely necessary conditions of optimality. If a feasible vector \underline{x} and a nonnegative vector \underline{y} happen to satisfy (3), (4), and (5), there is no guarantee that \underline{x} must be an optimal solution of (1). One cannot expect too much from first-order information.

Convexity of φ (equivalently, the positive semi-definiteness of \underline{H}) is second-order information of a type that guarantees the sufficiency of the Kuhn-Tucker conditions as an optimality criterion.

Theorem 2. Let $\varphi(\underline{x}) = \underline{c} \cdot \underline{x} + \frac{1}{2} \underline{x}^T \underline{H} \underline{x}$ be convex on \mathbb{R}^n . If \underline{x}^* is a feasible vector for (1) and there exists a vector \underline{y}^* such that $(\underline{x}^*, \underline{y}^*)$ satisfies (3), (4), (5), and (6), then \underline{x}^* is an optimal solution for (1).

In the convex QP case, the Kuhn-Tucker conditions are necessary and sufficient for optimality. The search for optimal solutions of such a problem can be confined to solutions of the Kuhn-Tucker system

$$\begin{aligned} \underline{u} &= \underline{c} + \underline{H} \underline{x} + \underline{A} \underline{x} \geq \underline{0}, \quad \underline{x} \geq \underline{0}, \quad \underline{x}^T \underline{u} = 0 \\ \underline{y} &= \underline{b} - \underline{A} \underline{x} \geq \underline{0}, \quad \underline{y} \geq \underline{0}, \quad \underline{y}^T \underline{u} = 0 \end{aligned} \quad (11)$$

3.2. Existence and Uniqueness. Nothing has been said so far about the existence and uniqueness of optimal solutions. One of the main results in this department is

Theorem 3. If the quadratic function Ψ is bounded below on the nonempty polyhedral convex set C , then Ψ attains its minimum on C .

To the author's knowledge, this result, known as the Frank-Wolfe theorem, first appeared in [24]. It does not require the convexity of Ψ , but when Ψ is strictly convex, it is possible to state a stronger theorem.

Theorem 4. A strictly convex QP with a nonempty constraint set has a unique optimal solution.

Although the uniqueness of the optimal solution is characteristic of all mathematical programs having strictly convex objective functions and convex constraint sets, the existence is not. It is apparent that for a strictly convex QP, the objective function $\Psi(\underline{x}) = \tilde{c} \underline{x} + \frac{1}{2} \underline{x}^T \underline{H} \underline{x}$ is bounded below on \mathbb{R}^n (and hence on C) by $\Psi(-\underline{H}^{-1} \underline{c}) = -\frac{1}{2} \underline{c}^T \underline{H}^{-1} \underline{c}$. Thus, the existence of the optimal solution is assured by the Frank-Wolfe theorem. More can and will be said about the existence of solutions in the sequel.

3.3. Duality. Associated with a convex QP (1) is a maximization problem called its dual. The dual is also a QP defined with the same data. It is customary to refer to (1) as the primal problem. Thus, for the primal problem

$$\begin{aligned} & \text{minimize } \Psi(\underline{x}) = \tilde{c} \underline{x} + \frac{1}{2} \underline{x}^T \underline{H} \underline{x} \\ & \text{subject to } \underline{A} \underline{x} \geq \underline{b} \\ & \quad \underline{x} \geq \underline{0} \end{aligned}$$

the corresponding dual problem is

$$\begin{aligned} & \text{maximize } \Psi(\underline{x}, \underline{y}) = -\tilde{b}^T \underline{y} - \frac{1}{2} \underline{y}^T \underline{A}^T \underline{H} \underline{x} \quad (12) \\ & \text{subject to } \underline{c} + \underline{H} \underline{x} + \underline{A} \underline{y} \geq \underline{0} \\ & \quad \underline{y} \geq \underline{0} \end{aligned}$$

This dual problem involves two vector variables: \underline{x} and \underline{y} . Only \underline{y} is required to be nonnegative (although it is known that nonnegativity can be imposed on \underline{x} without loss of generality).

Theorem 5. If \underline{x} is feasible for (1) and $(\underline{x}', \underline{y})$ is feasible for (12), then

$$\Psi(\underline{x}', \underline{y}) \leq \Psi(\underline{x}) \quad (13)$$

Moreover, if equality holds in (13), then \underline{x} and $(\underline{x}', \underline{y})$ are optimal solutions of (1) and (12), respectively.

This so-called weak duality theorem has several implications, one of which is that if the primal and its dual are both feasible, they must both possess optimal solutions--a consequence of the Frank-Wolfe theorem.

However, the main result in this direction is Dorn's duality theorem [21].

Theorem 6. If \underline{x}^* solves (1), then there exists a vector \underline{y}^* such that $(\underline{x}^*, \underline{y}^*)$ solves (12), and $\Psi(\underline{x}^*, \underline{y}^*) = \varphi(\underline{x}^*)$.

This can be seen by taking \underline{y}^* as the vector of Lagrange multipliers guaranteed by the Kuhn-Tucker theorem. By performing suitable manipulations, it is possible to verify the "converse duality theorem"

Theorem 7. If the program (12) has an optimal solution, then so does (1), and the optimal values of their respective objective functions are equal.

If a QP is feasible and fails to have an optimal solution its objective function must be unbounded in the direction of extremization. In fact, in the case of convex QP, the unboundedness of the objective function of one problem is equivalent to the infeasibility of its dual.

It should be noted that Dorn's duality theory of convex QP generalizes that of LP in nearly all respects. However, the structures of the paired QP problems are dissimilar. This is rectified in the symmetric duality theory developed by Cottle [8]. If one takes the primal as

$$\begin{aligned} \text{minimize } & \varphi(\underline{x}, \underline{y}) = \underline{c} \cdot \underline{x} + \frac{1}{2} \underline{x}^T \underline{H} \underline{x} + \frac{1}{2} \underline{y}^T \underline{Q} \underline{y} \\ \text{subject to } & \underline{A} \underline{x} - \underline{Q} \underline{y} \leq \underline{b} \\ & \underline{x} \geq \underline{0} \end{aligned} \quad (14)$$

then the dual is

$$\begin{aligned} \text{maximize } & \Psi(\underline{x}, \underline{y}) = -\underline{b} \cdot \underline{y} - \frac{1}{2} \underline{x}^T \underline{H} \underline{x} - \frac{1}{2} \underline{y}^T \underline{Q} \underline{y} \\ \text{subject to } & \underline{c} + \underline{H} \underline{x} + \underline{A} \underline{y} \geq \underline{0} \\ & \underline{y} \geq \underline{0} \end{aligned} \quad (15)$$

In these problems called symmetric dual quadratic programs, both \underline{H} and \underline{Q} are assumed to be positive semi-definite and symmetric. When $\underline{Q} = \underline{0}$, this pair of problems reduces to the dual pair considered by Dorn. The duality theorems quoted above are also valid for these symmetric dual programs.

Another result is called the Joint solution theorem.

Theorem 8. If (14) and (15) are both feasible programs, then there exists a pair of vectors $(\underline{x}^*, \underline{y}^*)$ which solves both problems, and in particular, $(\underline{x}^*, \underline{y}^*)$ yields a solution to the Kuhn-Tucker system (11).

4. QUADRATIC PROGRAMMING ALGORITHMS

4.1. Introduction. Among the many algorithms specifically designed for solving quadratic programs, three are preeminent. These are the methods of Beale [1], [2], [3], Wolfe [58], and van de Panne and Winston [47], [48], [50]. Each is a simplicial method in the sense that it involves a sequence of pivot steps similar to those of the simplex method of LP.

But, unlike the simplex method for LP, the pivot steps of these QP algorithms do not generate a sequence of adjacent extreme points of the primal constraint set \mathcal{C} . Instead, they work with a larger system of inequalities. In effect, they are concerned with satisfying the Kuhn-Tucker system (11) for the primal problem, though each does so in a different way. It is known that if the system (11) has a solution, then there exists a solution which is an extreme point of the set of points satisfying its inequality constraints:

$$\begin{aligned} \underline{u} &= \underline{c} + \tilde{H} \underline{x} + \tilde{A} \underline{y} \geq \underline{0} \\ \underline{y} &= \underline{b} - \underline{A} \underline{x} \geq \underline{0} \\ \underline{x} &\geq \underline{0} \\ \underline{y} &\geq \underline{0} \end{aligned} \tag{16}$$

These three QP algorithms are primal methods, i.e., they work with feasible solutions of the problem (1). Generally, they require an initialization step which obtains an extreme point of \mathcal{C} , or--what is the same thing--a basic feasible solution of the system

$$\begin{aligned} \underline{I} \underline{y} + \underline{A} \underline{x} &= \underline{b} \\ \underline{x} &\geq \underline{0}, \underline{y} \geq \underline{0} \end{aligned} \tag{17}$$

If the initialization step indicates the constraints are inconsistent, no further computation is required. Otherwise, a pivotal transformation of is obtained in which the vector on right-hand side of the equation is negative. Thus, the assumption that $\underline{b} \geq \underline{0}$ is tantamount to the assumption that an initialization procedure has already been applied and shown that the problem is feasible.

4.2. Beale's Method. One of the earliest and most successfully developed algorithms for QP is due to Beale [1], [2], [3]. The method proceeds from a basic solution of the linear constraints (17) expressed in the form

$$\begin{aligned} \underline{y} &= \underline{b} - \underline{A} \underline{x} \\ \underline{x} &\geq \underline{0}, \underline{y} \geq \underline{0} \end{aligned} \tag{18}$$

where $\underline{b} \geq \underline{0}$. This formulation makes it possible to express the objective function in terms of the nonbasic (independent) variables x_j . Furthermore, $\varphi(\underline{x}) = \tilde{c} \underline{x} + \frac{1}{2} \underline{x}^T \tilde{H} \underline{x}$ can be written in the form

$$\varphi(\underline{x}) = \frac{1}{2} \begin{bmatrix} 1 \\ \underline{x} \end{bmatrix} \begin{bmatrix} 0 & \tilde{c} \\ \underline{c} & \tilde{H} \end{bmatrix} \begin{bmatrix} 1 \\ \underline{x} \end{bmatrix}$$

At the basic feasible solution $(\underline{y}, \underline{x}) = (\underline{b}, \underline{0})$ the gradient of φ is given by \underline{c} , and it follows that if $\underline{b} > \underline{0}$ (nondegeneracy assumption) the value of φ can be decreased locally if and only if \underline{c} contains a negative component.

If, to be specific, $c_s = \min_j c_j < 0$, then an increase of the nonbasic variable x_s will bring about a decrease of the objective function value, at least locally. In terms of the initial data, the variation of the basic variables v_i is governed by the equation

$$v_i = b_i - a_{is} x_s \quad i = 1, \dots, m$$

The increase of the variable x_s is to be stopped as soon as one of the basic variables decreases to zero (as in LP) or the partial derivative $\partial \Psi / \partial x_s$ increases to zero. If neither of these occurs, then the objective function has no finite lower bound on \underline{c} and the procedure is terminated. If the increase of x_s is first blocked by v_i reaching zero, then a new basic solution is obtained as in the simplex method for LP. But, if the partial derivative vanishes first, then one tries to keep that partial derivative at the value zero (at least until such time as it is found advantageous to do otherwise). This is accomplished by introducing a variable

$$u_1 = c_s + \sum_{j=1}^n h_{sj} x_j \quad (19)$$

which is defined by the expression that gives the value of the partial derivative $\partial \Psi / \partial x_s$. This variable is said to be free in the sense that it is not sign restricted. In order to ensure that the partial derivative $\partial \Psi / \partial x_s$ remains zero, the equation (19) is adjoined to (18) thereby yielding $m + 1$ equations in $m + n + 1$ unknowns. In this enlarged system, the variable u_1 is made nonbasic in place of x_s (as in LP). As before, there are n nonbasic variables: u_1 and the x_j ($j \neq s$). The objective function can be represented in terms of the new set of nonbasic variables.

The procedure is repeated (with the new system) but with some special rules which come about due to the presence of free variables which by definition are allowed to negative as well as positive. If there is a possibility of decreasing the objective function by increasing or decreasing a free variable, this should be done--subject to the same principles that limit the variation of any nonbasic variable--before a sign-restricted variable is chosen for increase. When a free variable becomes basic, that variable and the corresponding equation are dropped from the problem because retaining them no longer serves a purpose.

Termination of the method occurs when either the nonbasic variable being increased (or decreased) from zero is not blocked by the vanishing of a basic variable or a partial derivative or else there is no longer a way to decrease the objective function by modifying a nonbasic variable.

This description of Beale's method is, of course, quite brief. In [1], [2], [3], [5], and [33] the reader will find many more details.

4.3. Wolfe's Simplex Method for QP. In [58], Wolfe showed how the simplex method of LP could be put to use in solving convex QP problems. His method can be regarded as a parametric convex QP procedure. (See Section 6 for more on this sort of problem.) It is designed to solve problems of the form

$$\begin{aligned} & \text{minimize } \Psi_{\lambda}(\underline{x}) = \lambda \underline{c} \underline{x} + \frac{1}{2} \underline{x}^T \underline{H} \underline{x} \\ & \text{subject to } \underline{A} \underline{x} \leq \underline{b} \\ & \quad \underline{x} \geq \underline{0} \end{aligned}$$

for all $\lambda \geq 0$. The matrix \underline{H} is assumed to be positive semi-definite; thus Ψ_{λ} is a convex function for all λ .

Actually, Wolfe gave two procedures which he called the short form and the long form. In the short form, λ is fixed, whereas in the long form it varies, thereby making it a parametric method.

The short form is valid for problems in which either \underline{H} is positive definite or the linear term $\lambda \underline{c} \underline{x}$ vanishes identically. The latter occurs when $\lambda = 0$ or $\underline{c} = \underline{0}$. It is initiated from a solution of the problem in which $\lambda = 0$ and to which the short form applies.

The Short Form. Consider the system

$$\begin{aligned} \underline{H} \underline{x} + \underline{A} \underline{y} - \underline{u} + \underline{z}' - \underline{z}'' &= -\lambda \underline{c} \\ \underline{A} \underline{x} + \underline{y} &= \underline{b} \\ \underline{x} \geq \underline{0}, \underline{y} \geq \underline{0}, \underline{u} \geq \underline{0}, \underline{v} \geq \underline{0}, \underline{z}' \geq \underline{0}, \underline{z}'' \geq \underline{0} & \end{aligned} \tag{20}$$

Initially, let \underline{x} , \underline{y} , and \underline{u} be zero. Select a basic feasible solution for the system (20) from among the variables in the vectors \underline{y} , \underline{z}' and \underline{z}'' . In particular, use the vector \underline{y} (it being assumed that $\underline{b} \geq \underline{0}$) and if $-\lambda c_j \geq 0$ use \underline{z}'_j ; otherwise, use \underline{z}''_j . Let \underline{z} denote the vector of n basic variables chosen from the \underline{z}'_j and \underline{z}''_j , and let \underline{E} denote the corresponding coefficient matrix. Now consider the LP

$$\begin{aligned} & \text{minimize } \underline{e}^T \underline{z} \\ & \text{subject to } \underline{H} \underline{x} + \underline{A} \underline{y} - \underline{u} + \underline{E} \underline{z} = -\lambda \underline{c} \\ & \quad \underline{A} \underline{x} + \underline{y} = \underline{b} \\ & \quad \underline{x} \geq \underline{0}, \underline{y} \geq \underline{0}, \underline{u} \geq \underline{0}, \underline{v} \geq \underline{0}, \underline{z} \geq \underline{0} \end{aligned}$$

where $\underline{e} = (1, \dots, 1)$. The algorithm calls for the solution of this LP by the simplex method with a restricted basis entry rule: if u_j is basic, do not allow x_j to be basic (and vice versa); if v_i is basic, do not allow y_i to be basic (and vice versa). The minimum value of the objective function (the sum of the artificial variables z_j) must be zero. An optimal solution of this LP obtained in the manner indicated will yield a solution of the Kuhn-Tucker conditions for the QP.

The Long Form. The possibility of solving the parametric problem

when $\lambda = 0$ is guaranteed by the short form of Wolfe's method. The solution of the parametric problem via the long form begins with the solution of the QP corresponding to $\lambda = 0$. Once this is in hand, one considers the LP

$$\begin{aligned} & \text{maximize } \mu \\ \text{subject to } & \tilde{H} \underline{x} + \tilde{A} \underline{y} - \underline{u} + \mu \underline{c} + \underline{E} \underline{z} = \underline{0} \\ & \underline{A} \underline{x} + \underline{y} = \underline{b} \\ & \underline{x} \geq \underline{0}, \underline{y} \geq \underline{0}, \underline{u} \geq \underline{0}, \underline{v} \geq \underline{0}, \underline{z} \geq \underline{0}, \mu \geq 0 \end{aligned}$$

and starts from the solution (given by the short form) in which $\tilde{x} \underline{u} = 0$, $\tilde{y} \underline{v} = 0$, $\underline{z} = \underline{0}$, $\mu = 0$. The maximization of μ is performed with the simplex method and the same restricted basis entry rules as in the short form. Furthermore, no \underline{z} -variable is permitted to become basic.

4.4. van de Panne and Winston's Simplex Method for QP. Still another prominent algorithm for convex QP is due to van de Panne and Winston [47], [48], [50]. Actually, these authors proposed several algorithms only one of which will be discussed here, namely their symmetric primal simplex method. A limited form of their "asymmetric method" was anticipated by Dantzig [18], [19].

The van de Panne-Winston symmetric primal simplex method provides an appropriate transition to the study of the linear complementarity problem. Accordingly, it is described in somewhat more detail than the methods of Beale and Wolfe. The method uses first-order information on the objective function φ to guide its steps. In this connection, the following observation is important:

Theorem 9. For any solution of the equations

$$\begin{aligned} \underline{u} &= \underline{c} + \tilde{H} \underline{x} + \tilde{A} \underline{y} \\ \underline{v} &= \underline{b} - \underline{A} \underline{x} \end{aligned}$$

it follows that

$$\tilde{c} \underline{x} + \frac{1}{2} \tilde{x} \tilde{H} \underline{x} = \frac{1}{2} (\tilde{c} \underline{x} - \tilde{b} \underline{y} + \tilde{x} \underline{u} + \tilde{y} \underline{v}). \quad (21)$$

The left-hand side of (21) is just the value of the quadratic objective function $\varphi(\underline{x})$. Theorem 9 has the further consequence that if $\tilde{x} \underline{u} = \tilde{y} \underline{v} = 0$, then the value of $\varphi(\underline{x})$ is given by the linear expression $\frac{1}{2} (\tilde{c} \underline{x} - \tilde{b} \underline{y})$.

In order to give a reasonable description of the algorithm, it will be helpful to consider the following change of notation. Let $\underline{x}_I = \underline{x}$, $\underline{x}_J = \underline{y}$, $\underline{y}_I = \underline{u}$, and $\underline{y}_J = \underline{v}$. The letters I and J stand for sets of indices. Also, let $\theta = 2 \varphi(\underline{x}_I) - \tilde{x}_I \tilde{y}_I - \tilde{y}_J \tilde{x}_J$. Initially $I = \{1, \dots, n\}$ and $J = \{n+1, \dots, n+m\}$, but these sets may change as the algorithm evolves. In the new notation, \underline{x}_I and \underline{x}_J are the primal variables. Those indexed by

I are nonbasic, while those indexed by J are basic.

In tabular form the relevant equations for the method are

	1	\underline{x}_I	\underline{y}_J	
θ	0	\tilde{c}	$-\tilde{b}$	
\underline{y}_I	\underline{c}	\underline{H}	\tilde{A}	
\underline{x}_J	\underline{b}	$-\underline{A}$	0	

(22)

The method consists of a sequence of pivot steps; each is one of the following three types:

- (i) An exchange of basic y_s and nonbasic x_s . (Index s is then transferred from I to J .)
- (ii) An exchange of basic x_t and nonbasic y_t . (Index t is then transferred from J to I .)
- (iii) An exchange of basic x_t and nonbasic x_s followed by an exchange of basic y_s and nonbasic y_t . (Index s is then transferred from I to J and index t is transferred from J to I .)

Pivot steps of these three types preserve a structural property of the tableau known as bisymmetry. Indeed, let the row corresponding to the objective function and the column of constants headed (headed by 1) be associated with the index 0. Then define $I_0 = \{0\} \cup I$. Next, regard the entries in the tableau as the elements of a matrix, say M . Then the principal submatrices $M_{I_0 I_0}$ and $M_{J J}$ are symmetric while $M_{I_0 J}$ is the negative transpose of $M_{J I_0}$. Such a combination of symmetry and skew-symmetry is called bisymmetry. The initial tableau (22) exhibits this property. An extremely important feature of the van de Panne-Whinston algorithm is that the pivot steps it uses lead to another bisymmetric tableau. More generally, one can state [29]

Theorem 10. Bisymmetry is preserved by principal pivoting.

Thus, at any stage of the algorithm, the current tableau can be written in the form

	1	\underline{x}_I	\underline{y}_J	
θ	κ	\tilde{c}'	$-\tilde{b}'$	
\underline{y}_I	\underline{c}'	\underline{H}'	\tilde{A}'	
\underline{x}_J	\underline{b}'	$-\underline{A}'$	\underline{Q}'	

(23)

where the index sets I and J in (23) may be different from those in (22).

The basic solutions encountered in the method all satisfy the

condition $\sum_I \underline{x}_I \underline{y}_I = \sum_J \underline{x}_J \underline{y}_J = 0$ so that $\theta = 2 \varphi(\underline{x}_I)$. The method is initiated with a basic feasible solution of the primal constraints. Thus, $\underline{x}_J = b \geq 0$ when $\underline{x}_I = 0$ and $\underline{y}_J = 0$. Hence, it remains to achieve the nonnegativity of \underline{y}_I .

Just as in the simplex method for LP, the van de Panne-Whinston algorithm entails the increase of one nonbasic variable at a time. This will be called the driving variable. In this algorithm, the driving variable is always an x -variable, say x_s , for which $y_s < 0$. The increase of the driving variable is blocked if and when the value of any basic variable belonging to a particular set reaches zero. In the present algorithm, this set consists of all basic x -variables (which are nonnegative) and one particular basic y -variable, y_s , whose value is negative. Preserving the nonnegativity of the x -variables means the same thing as preserving primal feasibility. The algorithm, then, goes as follows:

Step 1. Test for optimality. If $\underline{y}_I = \underline{c} \geq 0$, terminate. The current solution is optimal. Otherwise, let $c_s < 0$.

Step 2. If $H_{ss} = 0$, go to step 3. Otherwise, use x_s as the driving variable and determine the blocking variable.

- (a) If y_s is the blocking variable, perform a pivot of type (i), and return to step 1.
- (b) If x_t blocks x_s , and if $Q_{tt} > 0$, perform a pivot of type (ii) and repeat step 2 with x_s as the driving variable.

Step 3. Determine whether x_s is blocked. If $-A_{rs} \geq 0$ for all $r \in J$, terminate. The driving variable is unblocked and the objective function is unbounded below on the primal constraint set. Otherwise, let x_t block x_s .

- (a) If $Q_{tt} > 0$, perform a pivot of type (ii), and return to step 2 with x_s as the driving variable.
- (b) If $Q_{tt} = 0$, perform a pivot of type (iii), and return to step 1.

Each return to step 1 that does not lead to termination specifies a distinguished index $s \in I$ such that $c_s < 0$. Between such occurrences, the set of eligible blocking variables is composed of y_s and all the basic primal variables, i.e., those of the vector \underline{x}_J .

The reader may notice that no provision is made for the possibility of negative entries appearing as diagonal elements of H or Q . This can be justified by showing that the positive semi-definiteness of these matrices is preserved by the pivotal operations of this method. The general result along these lines is given in the next section. Keller [29] has extended this method to find a local minimum in a nonconvex QP.

5. THE LINEAR COMPLEMENTARITY PROBLEM

5.1. Formulation. Given a matrix $\underline{M} \in \mathbb{R}^{N \times N}$ and a vector $\underline{q} \in \mathbb{R}^N$, the linear complementarity problem $(\underline{q}, \underline{M})$ is that of finding a solution to the system

$$\underline{w} = \underline{q} + \underline{M} \underline{z} \quad (24)$$

$$\underline{w} \geq \underline{0}, \underline{z} \geq \underline{0} \quad (25)$$

$$\underline{z}' \underline{w} = 0 \quad (26)$$

(or showing that no solution exists). The LCP is stated in terms of $2N$ variables w_1, \dots, w_N and z_1, \dots, z_N . For each $i = 1, \dots, N$ the variables w_i and z_i constitute a complementary pair and each is the complement of the other. Conditions (25) and (26) imply that for each i

$$z_i w_i = 0 \quad (27)$$

This is the familiar complementary slackness property.

There is a nice geometrical interpretation of the LCP. If (24) and (25) have a solution, then \underline{q} belongs to the convex cone spanned by the nonnegative linear combinations of the columns of the matrix $[\underline{I}, -\underline{M}]$. The LCP asks to have \underline{q} so represented but in such a way that for all i , not both the i -th column of \underline{I} and the i -th column of $-\underline{M}$ are used. A matrix \underline{B} of order N such that for each i : $\underline{B}_i \in \{\underline{I}_i, -\underline{M}_i\}$ is called complementary with respect to \underline{M} . For a problem of order N , there are 2^N ways to select the matrix \underline{B} , although the choices need not all be distinct. The convex cones they span are called complementary cones. The problem $(\underline{q}, \underline{M})$ can then be construed as that of determining whether \underline{q} belongs to some complementary cones, and, if so, which one(s). An actual solution $\underline{w}, \underline{z}$ of the problem is given in terms of the weights used on the columns of a complementary matrix \underline{B} .

5.2. Examples. Linear complementarity problems arise in a number of ways, the most common being as Kuhn-Tucker systems (11) for quadratic programs. The formulation of (11) as a linear complementarity problem is accomplished by defining

$$\underline{M} = \begin{bmatrix} \underline{H} & \underline{A} \\ -\underline{A}' & \underline{0} \end{bmatrix} \quad \underline{q} = \begin{bmatrix} \underline{c} \\ \underline{b} \end{bmatrix} \quad \underline{w} = \begin{bmatrix} \underline{u} \\ \underline{v} \end{bmatrix} \quad \underline{z} = \begin{bmatrix} \underline{x} \\ \underline{y} \end{bmatrix}$$

An LCP of this type is special, if only by virtue of the bisymmetry of \underline{M} . In the convex QP case, the matrix \underline{M} (defined above) is positive semi-definite in the sense that

$$\underline{z}' \underline{M} \underline{z} \geq 0 \quad \text{for all } \underline{z} \in \mathbb{R}^N \quad (28)$$

This condition makes sense, even if \underline{M} is not symmetric. Moreover, it is very useful to look at positive semi-definiteness this way.

Not every LCP arises as a Kuhn-Tucker system, even though it may in fact be one. This is the case when the LCP represents equilibrium conditions for a mechanical (or economic) system [37], [30]. It might subsequently be discovered that these equilibrium conditions are the Kuhn-Tucker conditions for a QP. Under these conditions, alternate solution strategies may suggest themselves. That is, one may have a preference (possibly based on the availability of a software package) for solving the LCP as the QP whose Kuhn-Tucker conditions it represents.

However, this approach is not valid when the matrix \underline{M} lacks bisymmetry, and since there exist linear complementarity problems where this property is absent [13], [34], it is fortunate that the theory of linear complementarity does not rely on the bisymmetry condition. In this respect the LCP is more general than the QP.

5.3. Existence and Uniqueness Results. When a problem $(\underline{q}, \underline{M})$ is presented, it is very helpful to know what kind of a matrix \underline{M} is. Describing the full range of interesting possibilities would be inappropriate here, but the theorems quoted in this section cover the most important cases and should give an indication of how \underline{M} can influence the qualitative properties of the problem.

The matrix \underline{M} belongs to the class \underline{P} if and only if all principal minors of \underline{M} are positive, i.e.,

$$\det \underline{M}_{JJ} > 0 \quad \text{for all } J \subset \{1, \dots, N\} \quad (29)$$

Members of the class \underline{P} are called \underline{P} -matrices. Note that because \underline{M} is not assumed to be symmetric, condition (29) is not equivalent to positive definiteness. It is a fact, however, that if \underline{M} is positive definite in the sense that $\underline{z}^\top \underline{M} \underline{z} > 0$ for all $\underline{z} \neq \underline{0}$, then $\underline{M} \in \underline{P}$.

Theorem 11. The LCP $(\underline{q}, \underline{M})$ has a unique solution for every $\underline{q} \in \mathbb{R}^N$ if and only if $\underline{M} \in \underline{P}$.

This result, due essentially to Samelson, Thrall and Wesler [54], says (in the language of Section 5.1) that the complementary cones relative to \underline{M} actually partition \mathbb{R}^N .

A second result, one that is closely related to applications in engineering plasticity, concerns positive semi-definiteness. It was first proved by Cottle [9].

Theorem 12. If \underline{M} is positive semi-definite and the system (24), (25) has a solution, then $(\underline{q}, \underline{M})$ has a solution.

In the positive semi-definite case, neither the existence nor the uniqueness of a solution is guaranteed. However, when $(\underline{q}, \underline{M})$ is nondegen-

erate (that is, every solution of (24) has at least N nonzero components) and \underline{M} is positive semi-definite, then $(\underline{q}, \underline{M})$ has at most one solution [34].

The following is a general fact about linear complementarity problems that shows they have something more than linear constraints in common with LP problems.

Theorem 13. If the LCP $(\underline{q}, \underline{M})$ has a solution, it has a solution which is an extreme point of the set $Z(\underline{q}, \underline{M}) = \{(\underline{w}, \underline{z}) : \underline{w} = \underline{q} + \underline{M} \underline{z}, \underline{w} \geq \underline{0}, \underline{z} \geq \underline{0}\}$.

This observation has given rise to the notion of solving linear complementarity problems via linear programming techniques. This approach requires the a priori identification of a linear form that will be minimized at a complementary extreme point of the "feasible set." Although the subject is far too technical for inclusion in this paper, the reader is encouraged to consult [39], [40], [41], [15], [16] for a presentation of the idea and its significance. One simple instance where the LP approach is applicable warrants attention here.

The matrix $\underline{M} \in \mathbb{R}^{N \times N}$ belongs to the class \underline{Z} if and only if all of its off-diagonal entries (M_{ij} , $i \neq j$) are nonpositive. Elements of \underline{Z} are called \underline{Z} -matrices [23]. The following is due to Cottle and Veinott [17].

Theorem 14. If $\underline{M} \in \mathbb{R}^{N \times N}$, the following are equivalent:

- (i) for each $\underline{q} \in \mathbb{R}^N$, the set $Z(\underline{q}, \underline{M})$ contains a least element \underline{z}^* (that is, an element satisfying $\underline{z}^* \leq \underline{z}$ for all $\underline{z} \in Z(\underline{q}, \underline{M})$) and \underline{z}^* is the unique element of $Z(\underline{q}, \underline{M})$ satisfying $\underline{z}^* (\underline{q} + \underline{M} \underline{z}^*) = 0$;
- (ii) $\underline{M} \in \underline{P} \cap \underline{Z}$

Notice that whenever a polyhedral set \underline{C} contains a least element, that element minimizes any positive linear form of the set. In the case of an LCP $(\underline{q}, \underline{M})$ where $\underline{M} \in \underline{P} \cap \underline{Z}$, its unique solution is the least element of $Z(\underline{q}, \underline{M})$ and therefore solves the LP

$$\begin{aligned} & \text{minimize } \underline{c} \cdot \underline{z} \\ & \text{subject to } \underline{M} \underline{z} \geq -\underline{q} \\ & \quad \underline{z} \geq \underline{0} \end{aligned}$$

where $\underline{c} = (1, \dots, 1) \in \mathbb{R}^N$

5.4. Pivotal Algebra. The method described in Section 5.5 and others depends heavily on some algebraic results. One of these will be presented here in order to give more meaning to what is done in the sequel.

Consider the case where the ingredients of the equation (24) are partitioned as follows

$$\begin{aligned} \underline{w}_I &= \underline{q}_I + \underline{M}_{II} \underline{z}_I + \underline{M}_{IJ} \underline{z}_J \\ \underline{w}_J &= \underline{q}_J + \underline{M}_{JI} \underline{z}_I + \underline{M}_{JJ} \underline{z}_J \end{aligned}$$

The subscripts I and J represent complementary subsets of $\{1, \dots, N\}$. It is possible to solve for z_I in terms of w_I and z_J provided \underline{M}_{II}^{-1} exists. If it does, and the exchange is executed, the equivalent system

$$\underline{z}_I = -\underline{M}_{II}^{-1} \underline{q}_I + \underline{M}_{II}^{-1} \underline{w}_I - \underline{M}_{II}^{-1} \underline{M}_{IJ} \underline{z}_J$$

$$\underline{w}_J = \underline{q}_J - \underline{M}_{JI} \underline{M}_{II}^{-1} \underline{q}_I + \underline{M}_{JI} \underline{M}_{II}^{-1} \underline{w}_I + (\underline{M}_{JJ} - \underline{M}_{JI} \underline{M}_{II}^{-1} \underline{M}_{IJ}) \underline{z}_J$$

results. The coefficient matrix

$$\underline{M}' = \begin{bmatrix} \underline{M}_{II}^{-1} & -\underline{M}_{II}^{-1} \underline{M}_{IJ} \\ \underline{M}_{JI} \underline{M}_{II}^{-1} & \underline{M}_{JJ} - \underline{M}_{JI} \underline{M}_{II}^{-1} \underline{M}_{IJ} \end{bmatrix}$$

is called a principal transform of

$$\underline{M} = \begin{bmatrix} \underline{M}_{II} & \underline{M}_{IJ} \\ \underline{M}_{JI} & \underline{M}_{JJ} \end{bmatrix}$$

and the process is called principal transformation or principal pivoting.

The next theorem summarizes four separate results on principal pivoting.

Theorem 15. Let $\underline{M}' \in \mathbb{R}^{N \times N}$ be a principal transform of \underline{M} .

- (i) If \underline{M} is a P -matrix, so is \underline{M}' . See [57].
- (ii) If \underline{M} is positive definite, so is \underline{M}' . See [10].
- (iii) If \underline{M} is positive semi-definite, so is \underline{M}' . See [10].
- (iv) If \underline{M} is bisymmetric, so is \underline{M}' . See [29].

This theorem implies that when an algorithm calls for a principal transformation of the system (24), and \underline{M} has any of the properties mentioned in (i) - (iv), it will be preserved. Parts (iii) and (iv) of this theorem are used in justifying the van de Panne-Whinston algorithm presented in Section 4.4.

5.5. The Principal Pivoting Method. In the case where $\underline{M} \in P$, the LCP $(\underline{q}, \underline{M})$ can be solved for any $\underline{q} \in \mathbb{R}^N$. One particular method for doing this is called the principal pivoting method [20], [9], [12]. In contrast to the quadratic programming methods discussed in Section 4, it does not use an objective function to guide the pivot steps. Instead, it consists of a sequence of major cycles each of which is associated with a distinguished basic variable that happens to be negative in value. The major cycle, in turn, consists of a sequence of principal pivot steps that (i) prevent already nonnegative basic variables from becoming negative, and (ii) terminate when the distinguished variable increases to zero. The complement of the distinguished variable is nonbasic and acts as the driving variable

throughout the major cycle. The variables that are eligible to block its increase are the distinguished variable and all the nonnegative basic variables. This distinguishes the principal pivoting method from the van de Panne-Whinston which in some other respects is similar.

The system (24) can be represented in the tabular form

	1	z_1	z_N
w_1	q_1	M_{11}	M_{1N}
\vdots	\vdots	\vdots	\vdots
w_N	q_N	M_{N1}	M_{NN}

The algorithms can be stated as follows:

Step 0. Start with the basic solution $(w, z) = (q, 0)$.

Step 1. If $q \geq 0$, stop. The solution is at hand. Otherwise, select an index s such that $q_s < 0$. (This might, but need not, be chosen so that q_s is the smallest component of q .) Designate w_s as the distinguished variable and z_s as the driving variable.

Step 2. Determine which variable blocks the driving variable. This is done by calculating

$$-q_t/M_{ts} = \min \{ -q_s/M_{ss}, \min_i \{ -q_i/M_{is} : M_{is} < 0 \} \}.$$

Step 3. Perform a principal pivot. If $t = s$ (i.e., the blocking variable is the distinguished variable) then exchange w_s and z_s ; return to Step 1. If $t \neq s$, hold z_s at the value $-q_t/M_{ts}$ and perform a principal pivot, exchanging w_t and z_t . Return to Step 2.

It should be pointed out that some refinements are required to keep track of the true identity of the basic and nonbasic variables. As the algorithm is stated above, the basic variables are denoted generically as w 's and the nonbasic variables as z 's.

Another important theoretical point is that to be perfectly correct, it is necessary to consider the problem of degeneracy. This can be done by introducing suitable perturbations or lexicographic ordering. It can also be done by implementing certain tie-breaking rules. See [22].

There are other principal pivoting methods, and other classes of problems to which they apply--most notably, the positive semi-definite case. See [27], [12], [13], [44], [45].

5.6. Lemke's Method. Another, more robust, method for the LCP is due to Lemke [34], [35]. There are different versions of the same underlying idea but regrettably there is only room in this paper for one of them.

Consider the auxiliary LCP

$$\underline{w} = \underline{q} + \underline{e} z_0 + \underline{M} \underline{z} \quad (29)$$

$$\underline{w} \geq \underline{0}, \underline{z} \geq \underline{0}, z_0 \geq 0 \quad (30)$$

$$\sum \underline{w} = 0 \quad (31)$$

A solution of this system with $z_0 = 0$ is necessarily a solution of $(\underline{q}, \underline{M})$. If \underline{e} is a positive vector, then for $\underline{z} = \underline{0}$ and a suitably large value of z_0 , the vector $\underline{w} = \underline{q} + \underline{e} z_0 \geq \underline{0}$, so (29) and (31) hold. But if $z_0 > 0$ holds, then (24) does not. The Lemke method preserves properties (29), (30), and (31) and works by means of special pivot selection rules to achieve the condition $z_0 = 0$. The system (29) can be represented by the tableau

	1	z_0	z_1	...	z_N
w_1	q_1	e_1	M_{11}	...	M_{1N}
\vdots	\vdots	\vdots	\vdots	...	\vdots
w_N	q_N	e_N	M_{N1}	...	M_{NN}

Lemke's scheme goes as follows.

Step 0. Start with the basic solution $(\underline{w}, z_0, \underline{z}) = (\underline{q}, 0, \underline{0})$.

Step 1. If $q \geq \underline{0}$, stop. The solution is at hand. Otherwise, let the index s be defined by the condition

$$- q_s / e_s = \max_i \{ - q_i / e_i \}$$

Perform a pivot step making z_0 basic in place of w_s . (Now both w_s and z_s are nonbasic.) Designate z_s as the driving variable.

Step 2. Determine whether the driving variable is blocked, and if so, by which variable. This is done by checking whether the column vector (in the current tableau) associated with the driving variable z_s is nonnegative.

If it is, then z_s is unblocked, and the procedure is terminated. Otherwise the increase of z_s is blocked by some basic variable which decreases to zero first.

Step 3. Perform a change of basis. Exchange the (basic) blocking variable and the (nonbasic) driving variable. If the blocking variable was z_0 , stop. A solution is at hand. If not, let the complement of the blocking variable (now nonbasic) be the new driving variable. Return to Step 2.

Remarks similar to those of Section 5.5 must be made about the

need for handling degeneracy in this method. (See Eaves [22] on this subject.) But equally important is the question of what can be deduced from the lack of a blocking variable in Step 2. This occurrence triggers termination of the procedure (on a ray as it is often called), and it raises the question of what it means in terms of solvability of the original problem $(\underline{q}, \underline{M})$. Some insight can be derived from the following result [34].

Theorem 16. If the Lemke algorithm applied to $(\underline{q}, \underline{M})$ terminates on a ray, there exists a nonzero, nonnegative vector \underline{u} such that

$$u_i (\underline{M} \underline{u})_i \leq 0 \quad i = 1, \dots, N \quad (32)$$

In the case of \underline{P} -matrices, the system (32) cannot have a nonzero solution, and consequently Lemke's algorithm solves this class of problems. When \underline{M} is positive semi-definite (or, more generally, copositive-plus), termination on a ray implies that (24) and (25) have no solution. While it is known that Lemke's method can be applied to other classes of linear complementarity problems, it is doubtful that the full extent of its applicability has yet been determined. Investigations along these lines are to be found in [22], [36], [25], among others.

6. PARAMETRIC METHODS

6.1. Introduction. In a number of instances, one is interested in solving a QP or LCP for all values of certain parameters within a specified range. One such is exemplified by the long form of Wolfe's simplex method for QP considered in Section 4.3. The problem treated there entails a special type of variation of the linear term in the objective function. In the case of Wolfe's method, the parametrization is as much nested in the method for solving the nonparametric problem as it is in the formulation of a truly parametric problem, such as the so-called portfolio selection problem to which it is manifestly applicable. See Markowitz [42]. For applications of the parametric LCP, see Maier [38].

6.2. Parametric QP problems. The quadratic program (1) is specified by the data \underline{A} , \underline{b} , \underline{c} , \underline{H} . The most commonly encountered parametric versions of the QP problem can be formulated as follows. Let Λ be an interval of the real line. For all $\lambda \in \Lambda$,

$$\begin{aligned} & \text{minimize } (\underline{c} + \lambda \underline{c}') \underline{x} + \frac{1}{2} \underline{x}' \underline{H} \underline{x} \\ & \text{subject to } \underline{A} \underline{x} \leq \underline{b} + \lambda \underline{b}' \\ & \quad \underline{x} \geq 0 \end{aligned} \quad (33)$$

where \underline{c}' and \underline{b}' are given vectors. If $\underline{b}' = \underline{0}$ and $\underline{c}' \neq \underline{0}$, the problem has a parametric objective function, while if $\underline{b}' \neq \underline{0}$ and $\underline{c}' = \underline{0}$, the problem has a parametric right-hand side.

For a given value of $\lambda \in \Lambda$, there is a corresponding QP which may or may not be feasible, may or may not have an optimal solution. Since parametric programming methods generally proceed from an optimal solution corresponding to a particular parameter value, there is a little problem about getting started. But, in the convex case, this is not terribly difficult. To find a value of λ for which the corresponding QP has an optimal solution, it suffices to write down the conditions

$$\begin{aligned} \underline{c} + \underline{c}' \lambda + \underline{H} \underline{x} + \tilde{\underline{A}} \underline{y} &\geq \underline{0} \\ \underline{b} + \underline{b}' \lambda - \underline{A} \underline{x} &\geq \underline{0} \\ \underline{x} \geq \underline{0}, \quad \underline{y} \geq \underline{0}, \quad \lambda \in \Lambda \end{aligned} \tag{34}$$

When Λ is an interval, testing these inequalities for a solution is the same sort of problem one meets in LP. If a solution $(\underline{x}^0, \underline{y}^0, \lambda^0)$ is found, then for $\lambda = \lambda^0$, the corresponding QP has an optimal solution. If (34) has no solution, then there is no value of $\lambda \in \Lambda$ for which the corresponding QP has an optimal solution. This is a consequence of Theorem 8 (and of Theorem 12). Thus the procedure can be made to start from λ^0 .

Among the types of questions that parametric QP models attempt to answer are

- (i) How does the optimal value vary with λ ?
- (ii) How does the optimal solution vary with λ ?
- (iii) Over what range of values of λ are particular sets of variables positive? of zero?

Parametric QP of the type described above can be treated as a parametric LCP. For the sake of brevity, this approach will be taken here.

6.3. Parametric LCP Problems. Given $\underline{q} \in \underline{\mathbb{R}}^N$, $\underline{p} \in \underline{\mathbb{R}}^N$ and $\underline{M} \in \underline{\mathbb{R}}^{N \times N}$, the parametric LCP consists of solving the family of ordinary LCP problems $(\underline{q} + \lambda \underline{p}, \underline{M})$ for all $\lambda \in \Lambda$. The Kuhn-Tucker conditions of the parametric QP (33) are subsumed in this formulation.

In the discussion below, it will be assumed that \underline{M} is either a F -matrix or is positive semi-definite. In the former case, $(\underline{q} + \lambda \underline{p}, \underline{M})$ has a unique solution for each $\lambda \in \underline{\mathbb{R}}$, hence for each $\lambda \in \Lambda$. In the positive semi-definite case, the problem has a solution for every λ for which

$$\underline{q} + \lambda \underline{p} + \underline{M} \underline{z} \geq \underline{0}, \quad \underline{z} \geq \underline{0} \tag{35}$$

has a solution. Accordingly, in order to get started, one can solve the problem of determining a $\lambda \in \Lambda$ for which (35) has a solution. If there is

none, there is nothing further to be done.

Let λ^0 be a parameter value for which $(\underline{q} + \lambda^0 \underline{p}, \underline{M})$ has a solution. One can assume that $\lambda^0 = 0$. (Otherwise, the change of variable, $\theta = \lambda - \lambda^0$ gives rise to an equivalent problem $(\underline{q} + \lambda^0 \underline{p}) + \theta \underline{p}, \underline{M}$ which is solvable for $\theta = 0$.) Thus, with $\lambda^0 = 0$, the solution procedure for $(\underline{q}, \underline{M})$ yields a principal transform of the problem, say $(\underline{q}', \underline{M}')$, in which $\underline{q}' \geq \underline{0}$. Let \underline{p}' be the corresponding transformation of \underline{p} . Thus $(\underline{q} + \lambda \underline{p}, \underline{M})$ becomes $(\underline{q}' + \lambda \underline{p}', \underline{M}')$ under the action of the pivoting. The variation of λ can now begin.

Without loss of generality, one may assume the problem is to be solved for all $\lambda \geq 0$. If negative values of λ are to be considered, one may simply reverse the sign of \underline{p} , and proceed as indicated below. If Λ is a bounded interval, the increase of λ can be stopped at an appropriate point. Notice that if $\underline{p} \geq \underline{0}$, the problem is trivial, since then $\underline{q} + \lambda \underline{p}$ is nonnegative for all $\lambda \geq 0$.

An algorithm is therefore needed to solve the following problem. Given $\underline{q} \in \mathbb{R}^N$, $\underline{p} \in \mathbb{R}^N$, $\underline{M} \in \mathbb{R}^{N \times N}$ where $\min_i q_i \geq 0 > \min_i p_i$, solve the problem $(\underline{q} + \lambda \underline{p}, \underline{M})$ for all $\lambda \geq 0$.

Recall that by an earlier assumption, \underline{M} is either a P -matrix or a positive semi-definite matrix. In order to simplify the statement of the parametric LCP algorithm below, it will be assumed that $\underline{q} > \underline{0}$ and for each positive value of the parameter λ , at most one basic variable vanishes. It is possible to dispense with this nondegeneracy assumption as in [11].

Since $\underline{q} > \underline{0}$, it follows that $(\underline{w}, \underline{z}) = (\underline{q}, \underline{0})$ solves $(\underline{q} + 0 \underline{p}, \underline{M})$. That is, one starts with a nonnegative complementary solution of (24). In the parametric algorithm, one increases λ (much like a driving variable) and maintains a solution of the LCP $(\underline{q} + \lambda \underline{p}, \underline{M})$.

Step 0. Initialization. Start with $\lambda = 0$ and $(\underline{w}, \underline{z}) = (\underline{q}, \underline{0})$. Regard λ as the driving variable.

Step 1. Monotonicity check. If the driving variable is unblocked, stop. No change of basis is required for larger values of λ .

Step 2. Determination of the next critical value. Determine the index r by the condition

$$-q_r/p_r = \min_i \{ -q_i/p_i : p_i < 0 \}.$$

The driving variable λ is blocked by the r -th basic variable when it reaches the critical value $-q_r/p_r$.

Step 3. Change of basis.

3.1. If $M_{rr} = 0$, go to Step 3.2. ($M_{rr} > 0$.) Pivot on M_{rr} holding λ at the current critical value. Return to Step 1.

3.2. If $M_{ri} \leq 0$ for all i , stop; the problem has no solution for larger values of λ . Otherwise (since $M_{rr} = 0$, $M_{rl} > 0$ for some i and M is positive semi-definite) determine the index s by the condition

$$-\frac{1}{M_{sr}} (q_s - \frac{p_s}{p_r} q_r) = \min \left\{ -\frac{1}{M_{ir}} (q_i - \frac{p_i}{p_r} q_r) : M_{ir} < 0 \right\}.$$

Then perform the block pivot on the nonsingular matrix

$$\begin{bmatrix} M_{rr} & M_{rs} \\ M_{sr} & M_{ss} \end{bmatrix}$$

Return to Step 1.

There is a simple geometric interpretation of the parametric LCP and the algorithm stated above. The set of points of the form $q + \lambda p$ for $\lambda \geq 0$ is a half-line emanating from q (corresponding to $\lambda = 0$) and heading out of the nonnegative orthant in the direction p . The nonnegative orthant is a complementary cone. The problem is to maintain a representation of the general point of this half-line, i.e., $q + \lambda p$, as a member of a complementary cone. Changes of basis (i.e., representation) occur when for a critical value of λ the half-line meets the boundary of a complementary cone. Termination of the algorithm in Step 1 means no more boundaries will be met, whereas termination in Step 3.2 means that for any value of λ larger than the current critical value, there is no complementary cone containing $q + \lambda p$.

In some applications, it is desirable to know whether the components z_i in a solution are nondecreasing functions of λ . This is the case if and only if the coefficients of λ in the rows of the basic z_i are non-negative. It may also be of interest to know this information in advance of the actual computation [38]; this question is investigated in [10].

7. COMPUTATIONAL EXPERIENCE

Surprisingly little of a systematic nature is known about the relative merits of quadratic programming and linear complementarity methods and their implementations in software packages. Several explanations might be offered for this, not the least of which would be the sheer expense of the undertaking; the development of one truly professional calibre code is a costly project, and the comparison of methods is more so. Still further, there appears to be no body of recognized test problems to which would-be

experimenters can apply the various methods. But a few reports have appeared, and these are briefly described below.

An early effort at comparison was published by van de Panne and Whinston [49]. They considered a formulation of Beale's method with their own Simplex Method for QP (the asymmetric version) and concluded that the latter is superior in terms of iterations and storage. This prompted a spirited response from Beale [4] who drew attention to the more compact (practical) version of his method presented in [3]. Beale emphasized two other points. One is that the Hessian of the quadratic objective function has low rank relative to the number of variables, and the practical version of his method takes advantage of this fact. The second is that the practical implementation of his method is far more capable of handling problems with large numbers of variables.

At about the same time, Moore and Whinston [43] published a computational study comparing Wolfe's algorithm with those of the Dantzig-van de Panne-Whinston family. Wolfe's method seemed less attractive.

A study by Braitsch [7] analyzed the methods of Dantzig (van de Panne and Whinston), Beale, and Wolfe (in the original version and a modified one due to Braitsch) on small randomly generated problems. He pointed out the various advantages of the different methods already mentioned here and seemed to indicate the slight superiority of Dantzig's method over the others.

On the LCP side, some encouraging computational experience was reported by Ravindran [51] who experimented with a modification of Lemke's method on linear programming problems. Ravindran confirmed that this adaptation of Lemke's method is identical to the so-called self-dual parametric algorithm of Dantzig [19]. He also reported the computational superiority of this approach over the primal simplex method. Ravindran has published a computer routine for Lemke's method in its general form [52].

Another, quite sophisticated, version of Lemke's method has been developed by Tomlin [55], [56]. His code, called LCPL, maintains a basis inverse in product form and uses LU decomposition to reinvert it at a user-specified frequency. Computational experience privately reported by users of this program suggests that it performs quite satisfactorily. A small amount of computational experience with NULEMKE, the forerunner of LCPL, is recorded in [12]. To this may be added the following table which gives merely an indication of the program's performance on a set of 5 problems of very different types

Problem order N	Iterations	Time (IBM 370/168)
15	8	0.23 sec.
36	34	0.93 sec.
105	87	2.08 sec.
520	481	16.21 sec.
526	176	27.15 sec. (run in WATFIV)

It is important to realize that although some methods are (theoretically) applicable to all members of a broad class--e.g., convex QP problems--there are occasions when special methods are better suited for the solution of problems belonging to a particular subclass. Illustrations of this remark are to be found in the papers [12],[14],[15],[26],[28],[44], [45] where large, specially-structured problems are attacked by specially designed methods which could not be presented here for reasons of length.

ACKNOWLEDGEMENTS

This study was partially supported by the Air Force Office of Scientific Research (Contract F44620-74-C-0079) and by the Office of Naval Research (Contract N-00014-75-C-0267).

REFERENCES

11. BEALE, E.M.L., "On Minimizing a Convex Function Subject to Linear Inequalities", Journal of the Royal Statistical Society, Vol. 17B, 1955, p. 173.
2. BEALE, E.M.L., "On Quadratic Programming", Naval Research Logistics Quarterly, Vol. 6, 1959, p. 227.
3. BEALE, E.M.L., "Numerical Methods", Nonlinear Programming, edited by J. Abadie, North-Holland Publishing Co., 1967, p. 135.
4. BEALE, E.M.L., "Note on 'A Comparison of Two Methods of Quadratic Programming'", Operations Research, Vol. 14, 1966, p. 442.
5. BLUM, E., and OETTLI, W., Mathematische Programmierung, Springer-Verlag, 1975.
6. BOOT, J.C.G., Quadratic Programming, North-Holland Publishing Co., 1964.
7. BRAITSCH, R.J., "A Computer Comparison of Four Quadratic Programming Algorithms", Management Science, Vol. 18, 1972, p. 632.
8. COTTLE, R.W., "Symmetric Dual Quadratic Programs", Quarterly of Applied Mathematics, Vol. 21, 1963, p. 237.

9. COTTLE, R.W., "Note on a Fundamental Theorem in Quadratic Programming", Journal of the Society for Industrial and Applied Mathematics, Vol. 12, 1964, p. 605.
10. COTTLE, R.W., "The Principal Pivoting Method of Quadratic Programming", Mathematics of the Decision Sciences, Part 1, edited by G.B. Dantzig and A.F. Veinott, Jr., American Mathematical Society, 1968, p. 144.
11. COTTLE, R.W., "Monotone Solutions of the Parametric Linear Complementarity Problem", Mathematical Programming, Vol. 3, 1972, p. 210.
12. COTTLE, R.W., "Computational Experience with Large-Scale Linear Complementarity Problems", Fixed Points, edited by S. Karamardian, Academic Press, 1977, p. 281.
13. COTTLE, R.W., and DANTZIG, G.B., "Complementary Pivot Theory of Mathematical Programming", Linear Algebra and Its Applications, Vol. 1, 1968, p. 103.
14. COTTLE, R.W., and GOHEEN, M.S., "A Special Class of Large Quadratic Programs", Technical Report SOL 76-7, Department of Operations Research, Stanford University, 1976.
15. COTTLE, R.W., and PANG, J.S., "On Solving Linear Complementarity Problems as Linear Programs", Technical Report SOL 76-5, Department of Operations Research, Stanford University, 1976.
16. COTTLE, R.W., and PANG, J.S., "A Least Element Theory of Solving Linear Complementarity Problems as Linear Programs", Technical Report 76-30, Department of Operations Research, Stanford University, 1976
17. COTTLE, R.W., and VEINOTT, A.F., JR., "Polyhedral Sets Having a Least Element", Mathematical Programming, Vol. 3, 1972, p. 238.
18. DANTZIG, G.B., "Quadratic Programming: A Variant of the Wolfe-Markowitz Algorithms", Research Report 2, Operations Research Center, University of California, Berkeley, 1961.
19. DANTZIG, G.B., Linear Programming and Extensions, Princeton University Press, 1963.
20. DANTZIG, G.B., and COTTLE, R.W., "Positive (Semi-) Definite Programming", Nonlinear Programming, edited by J. Abadie, North-Holland Publishing Co., 1967, p. 57.
21. DORN, W.S., "Duality in Quadratic Programming", Quarterly of Applied Mathematics, Vol. 18, 1960, p. 155.
22. EAVES, B.C., "The Linear Complementarity Problem", Management Science, Vol. 17, 1971, p. 612.
23. FIEDLER, M., and PTAK, V., "On Matrices with Nonpositive Off-Diagonal Elements and Positive Principal Minors", Czechoslovak Journal of Mathematics, Vol. 12, 1962, p. 382.
24. FRANK, M. and WOLFE, P., "An Algorithm for Quadratic Programming", Naval Research Logistics Quarterly, Vol. 3, 1956, p. 95.
25. GARCIA, C.B., "Some Classes of Matrices in Linear Complementarity Theory", Mathematical Programming, Vol. 5, 1973, p. 299.
26. GOHEEN, M.S., "A Decomposition Principle for Single Constraint Quadratic Programs", Technical Report SOL 77-2, Department of Operations Research, Stanford University, 1977.
27. GRAVES, R.L., "A Principal Pivoting Simplex Algorithm for Linear and Quadratic Programming", Operations Research, Vol. 15, 1967, p. 482.

28. KANEKO, I. "A Linear Complementarity Problem with a N by $2N$ 'P'-Matrix", Technical Report 76-4, Department of Industrial Engineering, University of Wisconsin, Madison, 1976.
29. KELLER, E.L., "The General Quadratic Optimization Problem", Mathematical Programming, Vol. 5, 1973, p. 311.
30. KENNEDY, M., "An Economic Model of the World Oil Market", Bell Journal of Economics and Management Science, Vol. 5, 1974, p. 540.
31. KUHN, H.W., "Nonlinear Programming: A Historical View", Nonlinear Programming (SIAM-AMS Proceedings, Vol. 9), American Mathematical Society, 1976, p. 1.
32. KUHN, H.W., and TUCKER, A.W., "Nonlinear Programming", Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, edited by J. Neyman, University of California Press, 1951, p. 481.
33. KUNZI, H.P., and KRELLE, W., Nonlinear Programming, Blaisdell Publishing Company, 1966.
34. LEMKE, C.E., "Bimatrix Equilibrium Points and Mathematical Programming", Management Science, Vol. 11, 1965, p. 681.
35. LEMKE, C.E., "On Complementary Pivot Theory", Mathematics of the Decision Sciences, Part 1, edited by G.B. Dantzig and A.F. Veinott, Jr., American Mathematical Society, 1968, p. 95.
36. LEMKE, C.E., "Recent Results on Complementarity Problems", Nonlinear Programming, edited by J.B. Rosen, O.L. Mangasarian, and K. Ritter, Academic Press, 1970, p. 349.
37. MAIER, G., "A Matrix Structural Theory of Piecewise Linear Elasto-plasticity with Interacting Yield Planes", Meccanica, Vol. 5, 1970 p. 54.
38. MAIER, G., "Problem--On Parametric Linear Complementarity Problems", SIAM Review, Vol. 14, 1972, p. 364.
39. MANGASARIAN, O.L., "Linear Complementarity Problems Solvable by a Single Linear Program", Mathematical Programming, Vol. 10, 1976, p. 263.
40. MANGASARIAN, O.L., "Solution of Linear Complementarity Problems by Linear Programming", Numerical Analysis, Dundee 1975, edited by G. Watson, Springer-Verlag, 1976, p. 166.
41. MANGASARIAN, O.L., "Characterization of Linear Complementarity Problems as Linear Programs", Technical Report 271, Computer Sciences Department, University of Wisconsin, Madison, 1976.
42. MARKOWITZ, H.M., Portfolio Selection: Efficient Diversification of Investments, Wiley, 1959.
43. MOORE, J.H., and WHINSTON, A.B., "Experimental Methods in Quadratic Programming", Management Science, Vol. 13, 1966, p. 58.
44. PANG, J.S., "On a Class of Least-Element Complementarity Problems", Technical Report SOL 76-10, Department of Operations Research, Stanford University, 1976.
45. PANG, J.S., "Some New and Efficient Algorithms for Portfolio Analysis", Technical Report 1738, Mathematics Research Center, University of Wisconsin, Madison, 1977.

46. van de PANNE, C., Methods for Linear and Quadratic Programming, North-Holland Publishing Co., 1975.
47. van de PANNE, C., and WHINSTON, A.B., "The Simplex and Dual Method for Quadratic Programming", Operational Research Quarterly, Vol. 15, 1964, p. 355.
48. van de PANNE, C., and WHINSTON, A.B., "Simplicial Methods for Quadratic Programming", Naval Research Logistics Quarterly, Vol. 11, 1964, p. 273.
49. van de PANNE, C., and WHINSTON, A.B., "A Comparison of Two Methods for Quadratic Programming", Operations Research, Vol. 14, 1966, p. 422.
50. van de PANNE, C., and WHINSTON, A.B., "The Symmetric Formulation of the Simplex Method for Quadratic Programming", Econometrica, Vol. 37, 1969, p. 507.
51. RAVINDRAN, A., "Computational Aspects of Lemke's Complementary Algorithm Applied to Linear Programs", Opsearch, Vol. 7, 1970, p. 241.
52. RAVINDRAN, A., "A Computer Routine for Quadratic and Linear Programming Problems", Communications of the ACM, Vol. 15, 1972, p. 818.
53. RITTER, K., "Stationary Points of Quadratic Maximum-Problems", Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, Vol. 4, 1965, p. 149.
54. SAMELSON, H., THRALL, R.M., and WESLER, O., "A Partition Theorem for Euclidean N-Space", Proceedings of the American Mathematical Society, Vol. 9, 1958, p. 805.
55. TOMLIN, J.A., "User's Guide for LCPL", Technical Report SOL 76-16, Department of Operations Research, Stanford University, 1976.
56. TOMLIN, J.A., "Programmer's Guide to LCPL", Technical Report SOL 76-25.
57. TUCKER, A.W., "Principal Pivotal Transforms of Square Matrices", SIAM Review, Vol. 5, 1963, p. 305.
58. WOLFE, P., "The Simplex Method for Quadratic Programming", Econometrica, Vol. 27, 1959, p. 382.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)